
**ADULT LITERACY AND LIFESKILLS SURVEY
INTERNATIONAL PLANNING REPORT**

**For Discussion at the first meeting of National Study Managers
September 23-24, 1999
Luxembourg**

**STATISTICS CANADA AND THE UNITED STATES NATIONAL CENTER
FOR EDUCATION STATISTICS
WITH THE SUPPORT OF
THE EDUCATION STATISTICS SERVICES INSTITUTE**

TABLE OF CONTENTS

1. INTRODUCTION
2. BACKGROUND
3. OBJECTIVES
 - 3.1 Short Term Objectives
 - 3.2 Medium Term Objectives
 - 3.3 Long Term Objectives
4. DATA REQUIREMENTS
5. DESIGN PARAMETERS
 - 5.1 General Design Considerations
 - 5.2 Survey Coverage
 - 5.3 Other Design Considerations
 - 5.4 Sample Size Requirements
 - 5.5 Sample Allocation
 - 5.6 Sample Selection
 - 5.7 The Assessment Design
 - 5.8 Pilot Surveys
 - 5.9 ALL Quality Assurance Related to Sample and Assessment Design
 - 5.10 Revised Design Recommendations
6. SURVEY INSTRUMENTATION AND ADMINISTRATION
 - 6.1 Data Collection Instruments
 - 6.2 Data Collection Procedures
 - 6.3 Pilot Survey
 - 6.4 Data Collection for ALL
7. DATA PROCESSING
 - 7.1 Database Development
 - 7.2 Scoring

-
- 7.3 Statistical Analysis
 - 7.4 Data Processing for ALL
 - 7.5 Statistical Analysis for ALL

8. REPORTING AND DISSEMINATION

- 8.1 Data Products
- 8.2 Analytic Products
- 8.3 Reporting and Dissemination for ALL

9. QUALITY ASSURANCE CONSIDERATIONS

- 9.1 Introduction
- 9.2 Considerations for the National Planning Report
- 9.3 Team Composition
- 9.4 Expert Meetings
- 9.5 Survey Instruments
- 9.6 Sampling and Weighting
- 9.7 Data Collection
- 9.8 Data Capture, Editing and Coding
- 9.9 Test Scoring

10. MANAGEMENT, STAFFING AND SCHEDULE

- 10.1 International Management and Advisory Structure
- 10.1 National Management and Advisory Structure
- 10.2 Project Schedule, Tasks and Deliverables

11. BUDGET

- 11.1 The Item Pilot Phase
- 11.2 The Full Pilot and Main Data Collection Phases

12. NEXT STEPS

APPENDICES

APPENDIX A Roles and Responsibilities

APPENDIX B: International Partnerships and Framework Development Teams

EXECUTIVE SUMMARY

Several OECD countries have expressed an interest in fielding a comparative survey to measure the distribution in the adult population of a range of skills thought to be important to social and economic success.

This report proposes such a survey, which has been called the Adult Literacy and Lifeskills survey (ALL) (formerly known as the International Life Skills Survey (ILSS)).

Patterned on the International Adult Literacy Survey (IALS), the ALL would see the administration of direct performance tests to representative samples of adults aged 16 to 65.

A 30 booklet spiralled design will be employed to make optimal use of available testing time.

The study would directly assess performance in four skill domains:

- prose literacy

- document literacy

- numeracy

- problem-solving/analytic reasoning

The study would also assess two skills indirectly, through behavioural reports:

- Attitudes towards teamwork

- Access to and use of Information and communication technology

A background questionnaire would also be administered prior to testing.

Countries are asked:

- To name a national study manager

- To establish a national study team

- To draft a national planning report

- To contribute national expertise to the development of the item pools for prose literacy, document literacy and numeracy and to contribute to the development of the background questionnaire

To have the proposed framework documents for each skill domain reviewed by national experts

It is hoped that these activities will permit countries to take an informed decision regarding participation in the first round of data collection for the study by June, 2000.

Countries choosing to participate will be asked:

To field pilot study in the first quarter of 2001 to yield 1,200 cases.

To field a full survey, designed to yield 6,750 cases, in the first quarter of 2002.

An international comparative report will be released in December, 2003.

1. INTRODUCTION

The following report was drafted by the Special Surveys Division and the Center for Education Statistics of Statistics Canada, with the support of the Education Statistics Services Institute of the American Institutes for Research, and in cooperation with the Organization for Economic Cooperation and Development. The document is intended to inform the deliberations of national governments considering participation in the development and validation of the survey instruments for the Adult Literacy and Lifeskills survey (ALL). Any questions about the document should be addressed to:

T. Scott Murray
Director, CTCEs
Statistics Canada
R.H. Coats Bldg – 17 ‘A’
Tunney’s Pasture
Ottawa, Ontario K1A 0T6
Telephone: 613-951-9476
Facsimile: 613-951-0085
Internet: scotmur@statcan.ca

Marilyn Binkley
National Center for Education Statistics
555 New Jersey Ave., NW
Washington, DC 20208
U.S.A.
Telephone: 202-219-2195
Facsimile: 202-219-1736
Internet: marilyn_binkley@ed.gov

2. BACKGROUND

In 1994, nine countries, Canada, the US, the Netherlands, Switzerland, Poland, Germany, Sweden, France and Ireland fielded the International Adult Literacy Survey (IALS), the world's first large scale, comparative assessment of adult literacy. By December, 1995, Statistics Canada and the OECD had published *Literacy, Economy and Society: results of the first International Adult Literacy Survey* (OECD and Statistics Canada, 1995), a report which presented data for seven of the countries which participated in the first round of data collection.

Encouraged by the IALS success, five countries, Australia, New Zealand, Great Britain, Northern Ireland and the Flemish community in Belgium decided to administer the IALS instruments in 1996. Data from this round of collection was released in November, 1997 in *Literacy Skills for the Knowledge Society: Further results of the International Adult Literacy Survey* (OECD and HRDC, 1997).

Finally, nine countries, Chile, Finland, Norway, Denmark, Switzerland, Italy, Slovenia, the Czech Republic and Hungary, participated in a third, large scale round of data collection in 1998 that is scheduled to report in June, 2000.

Portugal, Japan and Malaysia have also successfully collected data with IALS instruments.

IALS provided hitherto unavailable information of the distribution of adult literacy skills and has provided tantalising insight into the causes and consequences of these skills for a range of countries. Key findings include:

- 1) important differences in literacy skills do exist across and within nations,
- 2) literacy skill deficits are found not just among marginalised groups, but affect large portions of the entire adult population,
- 3) literacy is strongly correlated with life chances and use of opportunities,
- 4) literacy is not synonymous with educational attainment,
- 5) literacy skills, like muscles, are maintained and strengthened through regular use, and
- 6) adults with low literacy skills do not usually acknowledge or recognize that their skills may pose a problem.

Not surprisingly, the IALS reports have attracted a great deal of interest from national policy makers and the popular press. It is clear that the study has answered many questions of pressing interest and concern. Yet, as with any well conceived study, IALS has raised as many questions as it has answered. Key among such questions are those which speculate about the relationship of literacy skill to other skills thought to be important to workforce productivity and labour market success. Many studies, including the Secretary's Commission About Necessary Skills (SCANS) in the United States, have posited the existence of a range of skill domains thought to be of economic importance. Yet little, if any, empirical evidence exists to test these notional skill frameworks.

The success of the IALS approach led several national governments to wonder if the methods could be adapted to measure a broader array of skills on an international level.

A first meeting to consider the possibility of mounting such a study was hosted by the Swedish Educational Authority Skolverket. Documentation prepared by Statistics Canada for consideration¹ reviewed the prevailing notions of basic skills and offered a hybrid typology with eight distinct domains which might be included in an IALS-type study. Statistics Canada suggested that the proposed study would, in each of the eight skill domains, administer a computer based test to a nested sample of workers within firms so that explicit statistical linkages would be available to isolate the impact of observed skill on productivity and profitability.

Interest in this idea was sufficient amongst national governments to organise a second meeting, this time hosted by the University of Amsterdam, to discuss the merits of available conceptual frameworks and to review the validity, reliability and operational feasibility of related measurement technology. This meeting concluded that coherent conceptual frameworks and satisfactory measurement technology did indeed exist for several, but not all, of the proposed skill domains. It became clear, however, that the costs and operational implications of fielding a computer-based test to a nested sample of workers within firms was well beyond the financial and technical capability of many of the prospective participants.

This realisation lead to a third meeting, this time of the International Study Team, to consider options. Hosted by the US National Center for Education Statistics in Washington DC, this meeting concluded, on pragmatic grounds, that:

- 1) the proposed assessment should use paper and pencil rather than computer-based tests,
- 2) only six skill domains should be measured: prose literacy, document literacy, numeracy, team work, problem solving, and working with information technology. While of interest, the other skill domains lacked either satisfactory theory or viable measures.

¹ See Background for Canadian Basic Job Skills Test, S. Jones, Human Resources Development Canada and Statistics Canada, 1996

3) the test should be administered to representative samples of adults drawn from households rather than from workers within firms.

These basic planning assumptions were subsequently ratified at both the first and second meetings of the ALL Project Advisory Committee (PAG), the Committee responsible for providing management oversight and advice to the Project within the OECD programme of work.

A fourth meeting of experts, convened in May, 1997, reviewed the proposed design of the study and the frameworks for each of the skill domains. Held at the OECD in Paris, this meeting concluded that additional work needed to be done on several of the skill frameworks if the study was to have a reasonable chance of generating valid, reliable and comparable skill profiles. As a result, the teams responsible for numeracy, problem solving, teamwork and practical cognition were funded to refine their frameworks and to collect sufficient empirical data to demonstrate the measurement properties of the proposed measures trans-nationally.

A fifth and sixth meeting of the International study Team were hosted by the U.S. National Center for Education Statistics in Washington, D.C. to review progress and plan for further development. Held in April 1998 and September, 1998, these meetings concluded that while the proposed frameworks for Problem Solving and Teamwork were judged to be adequate, the approach to measurement had failed to yield data of sufficient quality². In addition, the proposed instrumentation for measuring computer literacy was judged inadequate. As a result, new development teams were recruited and funded. Two additional meetings have been held. First, a meeting of all development team members was held in Washington, D.C. January 20 –23, 1999 to help integrate the different assessments and to provide expert feedback. A second meeting of development team leaders was held in Princeton, N.J. on August 23 –24, 1999 to come to final recommendations regarding inclusion in the ALL pilot as outlined below:

IBF in Germany were contracted to develop and test instrumentation for problem solving. A revised framework has been developed and a limited data collection carried out in Germany. This collection demonstrates that the approach to measurement is viable and it is proposed that IBF be commissioned to develop the balance of the items required for the pilot survey.

The American Institutes for Research were commissioned to develop and test instrumentation for teamwork skills. A limited data collection in Canada demonstrated that measures of attitudes to teamwork were valid and reliable but that behavioural measures of teamwork skill did not yield results of sufficient quality. ETS is continuing to work with the team to develop a scoring system that would yield better empirical results. It is recommended that the measures of attitudes to teamwork be included in the background questionnaire in the current round of ALL.

² see *Final Report for Validation of Problem Solving Measures*, National Center for Research on Evaluation, Standards and Student Testing, UCLA, 1998 and *Final Report for Validation of Teamwork Skills Questionnaire Using Computer-based Teamwork Simulations*, National Center for Research on Evaluation, Standards and Student Testing, UCLA, 1998.

The Ministère de l'Éducation Nationale in Luxembourg has agreed to invest in redeveloping the measures relating to information and communication technology. A first meeting of industry and measurement experts will take place immediately preceding the first meeting of ALL National Study Managers. Given the early stage of development of this component no recommendation can yet be made regarding inclusion. In the event that redeveloped ICT measures are not available in time to be included in the ALL pilot the International Study Team recommends that a revised version of the originally proposed questions on computer familiarity and use would be inserted in the background questionnaire. Prospective participants might be called upon to provide expert input.

An international team was formed to continue development of the ALL background questionnaire under the supervision of Westat and with input from the PEL sub-group of the OECD's INES Network B. The current draft of the questionnaire requires input from prospective participants to bring it to final form.

International teams were recruited to develop a roughly half of the items needed for the prose and document literacy domains. The balance of the item pools required for the ALL pilot remain to be developed by prospective participants.

The group of international experts responsible for development of the numeracy framework and instrumentation successfully collected empirical data in the United States and in the Netherlands. This data suggests that high quality comparative data can be collected in the first round of the ALL. The balance of the item pool required for the ALL pilot survey remain to be developed by prospective participants.

Limited empirical data for the practical cognition domain was collected in Spain and the United States. Although the results are intriguing reservations remain concerning the interpretability of the estimates flowing from the scoring system. The International Study Team recommends against the inclusion of measures of practical cognition in the core ALL study at this time.

In summary, the International Management Team believes that the ALL study, as currently conceived, will yield scientifically credible and policy relevant data. The project is deemed to be operationally feasible and financially viable given 12 countries chose to participate. The balance of this document outlines in detail the nature of the proposed study.

3. OBJECTIVES

The proposed study, which has been named the Adult Literacy and Lifeskills survey (ALL), has several sets of related objectives.

3.1 Short Term Objectives

The ALL has four short term objectives related to the development and validation of the ALL instruments, as outlined below:

First, the study hopes to:

receive feedback on the conceptual framework papers developed for each proposed skill domain

Framework documents for the ALL skill areas (prose and document literacy, numeracy, computer familiarity, teamwork skills, and analytic reasoning) have been made available for the first meeting of National Study Managers. Participants will be invited to circulate these documents among experts in their countries and to submit comments on them by December, 1999. This exercise is crucial to building national understanding of the scientific merits of the study.

Secondly, the study aims to:

establish an international consortium to develop and refine the measurement instruments

The IALS study used a collegial process involving most of the countries participating in the first round of collection to build the background questionnaire and the test instruments employed in the study. This process served a number of important purposes.

First, it served to provide participants with a fundamental appreciation of the conceptual framework and theory which underlay the test and how the theory manifested itself in measurement. This put national study teams in a much stronger position to explain and interpret their national findings.

Second, it provided the study with an international flavour which helped to establish the face validity of the measures and to counteract criticisms of cultural bias. Such criticisms have figured large in the IALS study.

Thirdly, it served to spread the costs associated with development across more actors. The development of direct assessments of the scale and complexity of IALS are an expensive undertaking. Sharing of development effort is an attractive way to keep overall costs down, a way which does not involve the physical transfer of money between countries.

Finally, the approach generated a sense of ownership and mutual trust amongst the participants. Ownership and trust is an important element in achieving high quality in such a complex and demanding study.

For these same reasons the ALL seeks to establish an international consortium to develop and refine the measurement instruments. National governments are invited to identify and finance national experts who could participate in the development of the balance of the required item pools for:

Prose literacy

Document literacy

Numeracy, and,

The background questionnaire.

Thirdly, the study would like to:

have up to fifteen countries field an item pilot survey in the first quarter of 2000

A pilot study is required to provide information on the psychometric performance of the test items for the purpose of selecting the best performing items for inclusion in the main study. As in IALS, the pilot also exposes countries to the operational exigencies of the study, allowing them to fine tune their cost estimates and quality assurance procedures. National governments are asked to identify a national study manager for the study. The immediate role of the national study manager would be to recruit a national study team for the purposes of developing a national planning report which in turn would be used to secure funding for the study and meet the quality assurance requirements of the project.

Finally, the study seeks to:

recover a share of the international overheads associated with the design and implementation

Statistics Canada and the US National Center for Education Statistics have collectively invested a considerable amount of time and effort in the development of this proposal and related documents such as the conceptual frameworks. Additional costs will be incurred in the management of the design of the item pools and their subsequent piloting. Statistics Canada proposes that each participant to the study contribute to help offset these costs.

In summary the overall goal of the ALL in the short term will be to develop and validate instruments for the assessment of analytic reasoning, team work, numeracy, prose literacy and document literacy on an international scale. The main product associated with this phase of the study will be a revised version of this report which will present the results of the piloting and which will recommend a final

design for the study. This report would be provided to the entire OECD membership for consideration.

3.2 Medium Term Objectives

The ALL has several medium term objectives that relate to the conduct of the main study in 2002. These objectives roughly parallel those set for the original IALS study.

First, the study hopes to:

profile the distribution of analytic reasoning, attitudes to team work, and numeracy in the adult population, determine the relationship of each of these skills, one to another, and to prose literacy and document literacy

The first two rounds of IALS data collection have already yielded an empirical data set of some 50,000 observations. With another 50,000 cases collected in the SIALS round, researchers will soon have sufficient sample size to generate a detailed understanding of the social and economic causes and consequences of the observed skill data. A basic objective of the ALL will be build on this understanding by adopting a design which will allow an empirical appreciation of how performance on each of the newly tested skills relates to the skills tested in the original IALS study. The rapid rate of technological innovation and the globalization of markets has lead to high rates of structural adjustment in many OECD economies. This, in turn, has quickened the pace at which disadvantaged individuals become marginalised. In many cases the very structure of educational systems and the labour market work against rapid adjustment and the interests of the marginal workers because the systems for signalling skill seek to divide the workforce into discrete, non-transferable categories.

The ALL seeks to empirically establish the existence of generic skill clusters which transcend industry, occupation, educational qualifications and age based experience.

determine the relationship of each of the tested skills to individual economic and social success

Many studies have documented the relationship of educational attainment to social and economic success but, until IALS, few studies had allowed one to explore empirically how this relationship depended, in turn, on more fundamental processes such as actual demonstrated skill. IALS has revealed that literacy skill and education are not synonymous and that social and economic success depend, in part, on tested skill. The study has also revealed interesting variation in these relationships both within occupations and between countries, facts that fit with common wisdom about the impact of economic and social organisation on the market for skill. The ALL will allow for an extension of this basic analysis to additional skill domains.

Identify sub-populations whose performance place them at risk

Much of the rhetoric employed in the North American debate about skills has been focussed on their

impact on the so called “high performance” workplace. Despite this fact, much of the attention of governments continues to be directed towards those groups of people whose skill levels place them at risk of being marginalised socially and economically. IALS has revealed that, in some economies at least, individuals with poor skills absorb significant wage and employment penalties. The design of policies and programmes to attenuate the worst of these impacts and to provide remedial education depend entirely on understanding the number, geographic distribution and characteristics of those so affected. The ALL will attempt to profile those whose performance place them at risk.

3.3 Long Term Objectives

The ALL hopes to meet a number of longer term objectives, including:

to shed light on the causes and consequences of the observed skill distributions

In order to truly understand the causes and consequences of any human phenomena, one needs longitudinal data, as it is only longitudinal data that allows one to disentangle cohort, life cycle and period effects. Thus the ALL, conceived to provide as cross-sectional “snapshot” of the distribution of skill in several domains, cannot be expected to advance our understanding in this regard. When analysed in conjunction with key covariates provided by the background questionnaire, however, the ALL can be expected to yield tantalising evidence about the relative impact that the various factors might have on the observed distributions of skill. This will be particularly so in those countries where IALS data has already been collected. In this case changes in the distribution of prose and document literacy skill between the two observations can be related in a synthetic longitudinal analysis, to changes in the underlying co-variables. Given the nascent state of skill measurement at the population level, the high cost of surveying longitudinally, the length of time it takes a longitudinal study to yield data and the horrendous cost of measuring the wrong things on such a study, the ALL can provide critical data to inform the next generation of true longitudinal studies.

to contribute to the literature on the basis of human cognition

The IALS study was based on a powerful theory of adult reading, a theory that served as a basis for both the design of the assessment and the analysis of the resultant data. Thus, IALS afforded researchers a unique opportunity to empirically validate the theory with data of unparalleled coverage and previously unknown sample sizes. The ALL offers a much greater potential to shed empirical light on the validity of several of the various competing frameworks which deal with the organisation and structure of human intelligence and cognition.³ A long term objective is, therefore, to assure that the ALL data is made available to researchers for this purpose.

³ See for example, Toward a triarchic theory of human intelligence, Robert J. Sternberg, 1984 Cambridge University Press, 1984 and Frames of Mind: The Theory of Multiple Intelligences, Howard Gardner, Harper Collins, 1993

to foster continued international co-operation on the design, implementation and analysis of data on the distribution and co-variates of skill

Direct measurement of the sort employed in the IALS study requires considerable operational and technical skill and significant financial resources to design, validate, collect and analyse. Such resources are beyond the means of many of even the most advanced economies to support. IALS has also demonstrated the potential of a comparative perspective to shed light on deep relationships underlying the observed phenomena, relationships that remain undetected in idiosyncratic national studies. Together, these facts argue strongly for the continuation of international collaboration in studies such as the Adult Literacy and Lifeskills survey, aided and abetted by the OECD and other multi-lateral agencies.

4. DATA REQUIREMENTS

The ALL is intended to support the development of thoughtful and fair policy responses to literacy and basic competency skill deficits observed in the adult population.

The data requirements that flow from this objective are not, however, obvious. To begin with it must be acknowledged that, relative to school-based tests or standard household-based surveys, such studies are expensive. In Canada, the average cost per unit for IALS was \$160 US⁴. Faced with this magnitude of collection costs, the challenge is to select a sample that one can afford and which is sufficiently large to provide enough cases to support analysis across obvious covariates. For ALL, a minimum sample yield of 6,300 is set. This size must be inflated to account for expected non-response and any design effect associated with the sample design used.

A little information is warranted to put this sample size in perspective. Statistically speaking roughly 600 completed cases are required to support a reliable profile of literacy and basic competency ability for any given population. Thus, assuming population subgroups of equal size, a 6,300 sample size might support estimates for up to ten subgroups and, within each such subgroup, one could be able to estimate any characteristic which represents a minimum of 15% of the subgroup. In addition, it is important to note that the statistical methodology employed to derive the estimates of ability demand a minimum of 1,000 completed cases to yield stable parameters.

It is proposed that ALL participants be required to field samples which will yield the equivalent of 6,300 randomly selected completed cases. Assuming a 70% response rate, countries are obliged to field roughly 9,650 cases to achieve this goal. Experience with IALS suggests this to be a reasonable

⁴ In IALS collection costs per unit varied greatly from country to country. From a low of US \$30 in Poland to a high of US \$330 in the US. Such variance is well beyond that which can be explained by differences in prevailing wage rates. Prospective participants are encouraged to employ the basic design parameters presented in this document to arrive at an estimate for their own situation.

target. Samples larger than this are to be encouraged, particularly if there is interest in looking at particular population subgroups. The addition of relatively rare sub-populations must, however, be accomplished with care as their inclusion can serve to reduce the statistical efficiency of the sample for the general population. Statistics Canada will, through the ALL quality assurance process, offer advice and guidance on this issue.

5. DESIGN PARAMETERS

Chapters 5, 6 and 7 of this report deal with substantive aspects of the design to be followed in ALL. Many of the design parameters spelled out in these three chapters are adapted from the successful IALS study. They describe the design parameters for the ALL main survey, planned for early 2002, as well as for a full pilot, planned for the first quarter of 2001. Additional details are provided in Quality Assurance Specifications for the Adult Literacy and Lifeskills survey that will be distributed at the first meeting of ALL NSM's.

5.1 General Design Considerations

Sample design options for ALL-type assessments are principally determined in consideration of detailed data requirements and the need to carry out lengthy assessments in respondents' homes. The detailed data requirements set out earlier imply the selection of a sample that represents both the population of interest and the ability distribution.

This latter requirement favours sample frames that can provide information, such as educational attainment and age, which is known to be highly correlated with literacy and basic competency ability. This type of information allows for the available sample to be allocated in such a way to maximize the information yield of the assessment.

The second consideration, the need to travel to respondents' homes to administer the assessment, favours frames which incorporate some degree of clustering, although care must be taken not to unduly increase the design effect in so doing.

Both of these general design considerations are, for example, met by vehicles such as the EU's Community Labour Force Survey.

National Study Teams will be required to nominate a survey statistician to participate in two design development meetings – one in advance of submitting the National Planning Report and the second after the conduct of the pilot survey but prior to finalising the statistical design of the main survey.

5.2 Survey Coverage

Given the objectives and data requirements set out above it is perhaps a statement of the obvious that the sample should seek to represent the entire adult population 16-65 years of age (inclusive). Coverage exclusions should not exceed 5% of the population in this age range for the geographic area to be represented.

By way of example, the Canadian component of IALS excluded, for a variety of methodological and operational reasons, inmates of institutions, full time members of the armed forces and residents of Indian Reservations, the Yukon and Northwest Territories. Collectively these exclusions represent slightly less than 2% of the adult population.

Countries are free to include individuals older than 65 but, in doing so, must not reduce the sample allocated to the 16-65 year old age group. Older adults also tend to have lower response rates, thereby increasing survey costs. Issues such as these will need to be judged relative to the information that may be gained.

5.3 Other Design Considerations

In light of both the response burden and the likely positive intra-class correlation of literacy and basic competency ability among members of the same household, it is deemed appropriate to limit the selection of respondents to one person per household.

5.4 Sample Size Requirements

The sample size for ALL was based upon a balanced consideration of the major data requirements, the level of precision desired for estimates produced at various levels of aggregation, and an appreciation of the collection cost associated with each unit.

Although the survey is intended to produce a wide range of estimates, the following two estimates were deemed to be the most important for the purpose of determining appropriate sample sizes for the study:

- (1) a national literacy and basic competency profile
- (2) a literacy and basic competency profile for each important sub-population

i. Sampling Precision

The coefficient of variation (CV) is a measure commonly used to express the precision of survey estimates. A relative measure of precision, it relates the standard error of an estimate to its size. Specifically, the CV of an estimate is the ratio of the standard error of the estimate and the population value. Algebraically, the CV can be expressed as follows:

$$CV^2 = \frac{DEff (1 - p)}{n_0 r p}$$

where:

n_0 = sample size⁵

p = proportion of the population belonging to a literacy profile category.

$DEff$ = the design effect (since the above formulation is based upon a simple random sample, a factor is introduced to account for the relative efficiency of the sample design used as compared to a simple random sample).

r = anticipated response rate

Solving for n_0 , we have

$$n_0 = DEff \frac{(1 - p)}{CV^2 r p}$$

⁵ in light of its negligible impact in a population the size of Canada, a term to account for sampling from a finite population has been dropped for simplicity. Countries with smaller populations may wish to incorporate a finite population correction since it will serve to reduce the total sample marginally.

In order to determine the required sample size, one can specify the desired level of reliability to be associated with major survey estimates. Statistics Canada's release policy highlights three levels of reliability: estimates whose CV's below 16.5% or below can be released without qualification; estimates with CV's between 16.6% and 33.3% require some cautionary note as to their relative unreliability; estimates with CV's of over 33.3% are recommended not be released. To illustrate the use of CV's, an estimate with a CV of 16.5% is likely to be within 16.5% of the true value with 68% confidence, and within 33% of the true value with 95% confidence.

From the formula for N_0 above, it is evident that the smaller p is, the larger the sample size must be in order to attain a specified level of reliability. Therefore, measurement of rare characteristics or estimates of very fine breakdowns of the population require larger sample sizes. In order to determine an appropriate sample size for the breakdowns required, the following assumptions have been made:

- (1) the minimum proportion (min p) of the population which belong to any literacy profile category for which estimates are to be published is between .10 and .15.
- 2) The design effect used to account for the use of a complex multi-stage sample design is approximately 2. Higher design effects may result if rare sub-populations are oversampled.
- 3) The anticipated response rate is 70%.
- 4) The CV was set at 16.5%, the Statistics Canada standard required for the publication of estimates without qualification.

The smallest sample size required in each cell (breakdown) for which a literacy and basic competency profile is to be estimated is:

- (1) 600 in the case of min $p = 0.15$ and
- (2) 950 in the case of min $p = 0.10$.

5.5 Sample Allocation

As mentioned above, sample designs that incorporate individual characteristics are favoured because they allow for the available sample to be allocated efficiently for the purposes of estimating the ability distribution of the population.

For example, a country might find a relatively small proportion of the population with less than high school education and a relatively large proportion of the population with a college diploma or a university degree. Using available information to stratify the sample in advance allows more of the available sample to be devoted to the smaller poorly educated population, where literacy and basic competency abilities are likely to be less predictable, at the expense of the larger, well educated population where test performance will be highly predictable.

Although it will not be possible to produce statistically reliable results for small sub-groups of the population, it will be possible to carry out multi-variate analyses (correlation, regression, etc.) to gain

a better appreciation of the factors (eg. background variables) which influence literacy proficiency.

Statistics Canada has employed these techniques to successfully generate reliable estimates for small areas without sufficient sample sizes to support direct estimates.

5.6 Sample Selection

In keeping with the foregoing description, sample selection for ALL should be straight forward. By way of example, one could first select one person 16-65 at random from each household. The individual records would next be sorted according to age, educational attainment, etc. Within these categories, the file would be sorted further according to geographic strata in order to make the sample representative on a geographic basis. The required sample would then be systematically selected within these cells according to the targeted allocations. Where prior information is not available, similar results could be obtained by employing a "KISH" type sample selection label to impose differential probabilities of selection at the time of interview.

5.7 The Assessment Design

The ALL will include two skill areas previously assessed by the IALS, and two areas that are new to international assessments of adult skills. The two areas repeated from IALS are prose literacy and document literacy. The new areas are numeracy and analytic reasoning skills.

Prose literacy focuses on the knowledge and skills needed to understand and use information from texts that contain extended prose organised in a typical paragraph structure found in materials such as editorials, news stories, brochures and pamphlets, manuals, and fiction.

Document literacy focuses on knowledge and skills required to locate and use information found in qualitatively different printed materials that contain more abbreviated language and use a variety of structural devices to convey meaning. These include tables, charts, graphs, indices, diagrams, maps, and schematics.

Numeracy refers to the ability to interpret, apply, and communicate mathematical information in commonly encountered situations (adapted from Queensland Department of Education, 1994). Numeracy tasks can be characterised by the computational skills required, problem-solving strategies used.

The inclusion of *computer familiarity and use of information and communication technology* in the assessment recognises that the ability to use computer technology is becoming a necessity for participation in the labour market beyond the most rudimentary level jobs. For the ALL, computer familiarity includes access to computers, self-assessment of computer-related attitude and ability, use of and experience with computers, and use of and experience with related technology.

Analytic reasoning is a sub-domain of broader *Problem-solving* skill. *Analytic reasoning* involves the ability to apply cognitive processes toward determining a solution when that solution is not

immediately obvious to the problem solver. The ALL definition of analytic reasoning includes three elements: content understanding, meta-cognition (the conscious and periodic self-checking of whether one's goal is achieved and, when necessary, selecting and applying different strategies), and problem-solving strategies (which may be domain specific or domain independent).

Attitudes to teamwork influence how effective an individual can function as a member of a group.

A taxonomy of six teamwork processes was proposed for ALL. These are adaptability, recognising problems and responding appropriately; communication, the exchange of clear and accurate information; co-ordination, organising team activities to complete a task on time. ALL investments in measurement have been unable to provide items of sufficient quality to support valid and reliable scales. Only the attitudinal items meet the strict ALL criteria for inclusion.

The two literacy skills, prose and document, as well as numeracy and analytic reasoning, will be measured through cognitive assessments. Attitudes to teamwork and use of information and communication technology will be measured from behavioural reports via survey questions incorporated in the background questionnaire. Each survey respondent will be administered a background questionnaire, and a paper-and-pencil test. This will be further developed in the next chapter.

In order to ensure broad coverage of the skill domains and not to overburden individual assessment takers, a Balanced Incomplete Block (BIB) assessment design is proposed for ALL. With a BIB design, a large set of assessment tasks are organized into smaller sets of tasks, called blocks, each containing items from each of the skill domains that cover a wide range of difficulty i.e. from easy to difficult. Individual respondents are not required to take the entire set of tasks. Instead, randomly equivalent samples of respondents were to be administered one of 16 test booklets in the original ALL design. Each booklet would contain one of the two literacy skills and one of either numeracy or problem-solving, as shown below:

Booklet	Literacy		New Skill	
	Domain	Items	Domain	Items
1	Prose A	15	Numeracy A	30
2	Prose A	15	Numeracy B	30
3	Prose A	15	Problem Solving A	30
4	Prose A	15	Problem Solving B	30
5	Prose B	15	Numeracy A	30
6	Prose B	15	Numeracy B	30
7	Prose B	15	Problem Solving A	30
8	Prose B	15	Problem Solving B	30
9	Document A	15	Numeracy A	30
10	Document A	15	Numeracy B	30
11	Document A	15	Problem Solving A	30
12	Document A	15	Problem Solving B	30
13	Document B	15	Numeracy A	30
14	Document B	15	Numeracy B	30
15	Document B	15	Problem Solving A	30
16	Document B	15	Problem Solving B	30

In this design a total of 200 cases are required for each booklet in order to meet our sample size requirement of 6,300 cases for the main survey.

5.8 Pilot Survey

A full pilot survey, planned for the first quarter of 2001, will be required of all participants. This pilot is to be considered as a full dress rehearsal for the main data collection. This activity is described further in section 6.3. As originally proposed the required sample yield for the full pilot was 1,275 cases.

5.9 ALL Quality Assurance Related to Sample and Assessment Data

In order to contain risk associated with deviation from the prescribed design parameters, prospective ALL participants will be required to submit a detailed Planning Report which documents all relevant details on the proposed national sample design for:

- (1) the full pilot survey,
- (2) the main survey.

Participants will also be required to submit their final booklet variants for inspection. Statistics

Canada will provide a sample planning report based on the Canadian IALS study. Participants will be required to provide a data file detailing the proposed distribution of the sample for the main survey that will eventually be compared with actual yields. Statistics Canada will provide guidelines for these files.

5.10 Revised Design Recommendations

ETS was asked to review the proposed design of the assessment in light of the the results of the meeting of the development team leaders in August 23 - 25, 1999 in Princeton, NJ. This meeting brought together team leaders from each of the skill areas under consideration for inclusion in the ALL. Team leaders reported on the progress each team had made on further developing and refining their frameworks, item pools and, in some cases, conducting a feasibility study. As a result of the information presented, the ALL management team recommends that certain changes be made in the design for both the pilot and main assessments. The proposed changes outlined below reflect the following recommendations:

- Expand both the numeracy and problem solving pilot tests to result in testing four blocks of numeracy items and five blocks of problem solving items.
- Delete the one block of practical cognition and two blocks of teamwork items from the pilot design.
- Expand the main assessment design so that there will be four problem-solving blocks and a full covariance matrix among the literacy, numeracy and problem-solving domains.
- Revise the design in such a way that the total sample size does not extend beyond 7,000 respondents.

Revised Design for the ALL Pilot

The design for the pilot assessment assumes that the primary goal is to pilot test new items for each skill area and to establish from these pools new scales in each of the content areas that would have scale invariance across countries. No attempt will be made to

establish relationships among the scales or to report on outcomes as a result of the pilot data.

Once the scales are created, the main assessment would be designed to provide reliable and valid population estimates in each content area for each country and to estimate the interrelationships among the scales.

The pilot and main assessment designs now reflect the fact that both practical cognition and teamwork are no longer included in the cognitive phase of the survey. While there continues to be interest in these domains, they will have to be part of the background questionnaire or incorporated in some other manner if they are to be included in the ALL. The same is true for Information and Computer Technology (ICT) which is also of considerable interest to the PAG. Again, the ALL management team did not feel that the assessment framework was sufficiently developed to incorporate this domain into the cognitive phase of the survey design.

Developing the Pilot Test Design

The pilot test design being recommended for the ALL is relatively limited in scope to minimize costs and time needed to collect and analyze the data. This design will maximize the information that is needed to select the best items from among the various pools to create reliable and valid scales which are comparable across cultures and languages. This information will include:

- (1) traditional item statistics such as p-values and item reliabilities;
- (2) information on operational problems and scoring rubrics that may need adjustment or redevelopment;
- (3) translation issues that may affect item comparability;
- (4) timing data; and,
- (5) initial estimates of the invariance of items and scales across countries.

The design will not allow for the estimation of inter-scale covariances, for stable item parameters or for estimating the distributions of performance on the new scales.

It has been assumed the following with respect to each of the content areas:

- Numeracy will test four 30-minute blocks of approximately 20 items each.
- Problem Solving will test five 30-minute blocks where each block contains one project.
- Prose and Document Literacy will test seven 30-minute blocks of approximately 20 items each. These blocks will contain 100 new and 40 old items for equating purposes.

Revised Pilot Test Design for Life Skills

Literacy							Numeracy				Problem Solving				
L1	L2	L3	L4	L5	L6	L7	N1	N2	N3	N4	PS 1	PS 2	PS 3	PS 4	PS 5
1	2														
	2	3													
		3	4												
			4	5											
				5	6										
					6	7									
1						7									
							1	2							
								2	3						
									3	4					
							1			4					
											1	2			
												2	3		
													3	4	
														4	5
											1				5

Each line of the table represents a single field-test booklet consisting of 2 blocks of tasks. Each booklet needs to be administered to only 75 adults in each country since each block appears twice in this design and each item needs to be attempted by 150 respondents. This means that the sample size per country will need to be at least 1200 completed cases or 75 less than under the original recommended design.

Revised Design for the Main Survey

The main assessment will benefit from the pilot test design in that items will have been selected to construct reliable and valid scales which are invariant across cultures and language

groups thus reducing the number of items to be included in the main assessment. Thus we anticipate for the main design having two 30-minute blocks of Numeracy, four 30-minute blocks of Problem Solving and four 30-minute blocks of prose and document literacy items.

This design will allow us to report on the distributions of the Life Skills for each of the participating countries as well as compare among them using IRT rather than traditional item-scale statistics. In addition, it will provide a full covariance matrix among the literacy, numeracy and problem solving scales thus allowing us to estimate the discriminant validity among these scales.

Revised Design for the Main Assessment

Literacy					Numeracy		Problem Solving			
	L1	L2	L3	L4	N1	N2	PS1	PS2	PS3	PS4
L1		X			X		X			
L2			X			X			X	
L3				X		X		X		
L4	X				X					X
N1		X	X			X	X			
N2	X			X	X			X		
PS1		X						X		
PS2				X					X	
PS3	X				X					X
PS4			X			X	X			

Each “X” in the table above represents a two-block, 60-minute booklet that will be administered to a random sample of at least 225 adults within each country. Each booklet consists of a content block taken from the row and a content block taken from the column indicated. The first block in each booklet represents the content block indicated by the row. There are a total of 30 assessment booklets yielding a minimum sample for each country of 6,750 adults. It is important to note that this design doubles the amount of information that will be obtained for problem solving over the design originally proposed.

In this design, each problem-solving item appears 5 times, each literacy item 6 times and each numeracy item 8 times. This means that some 1125 adults will respond to each problem-solving item, some 1350 adults will respond to each of the literacy items and approximately 1800 adults will respond to each of the numeracy items. This information will be used to obtain stable item parameters within each country. It is also important to note that this design provides adequate coverage for estimating the covariance structure among the domains. For example, there are 8 blocks (or some 1800 cases within each country) that will contribute to the estimate of the covariance between literacy and numeracy as well as between literacy and problem solving. In addition, there are 4 blocks or some 900 cases that will contribute to the covariance estimate between numeracy and problem solving.

6 SURVEY INSTRUMENTATION AND ADMINISTRATION

6.1 Data Collection Instruments

Following the IALS model, ALL will employ of a number of distinct collection components in the main survey. These components include a screening questionnaire, a background questionnaire, and a set of simulation tasks.

i Screening Questionnaire

Where necessary, the screening questionnaire is administered in advance of the main survey to provide stratification variables that are not currently available from the selected sample frame.

ii Background Questionnaire

The background questionnaire serves a number of analytic purposes in ALL. First, it allows the identification of groups in each society whose literacy skill levels place them at the highest risk and can reveal the associated risk factors. Second, the questionnaire identifies perceived barriers and/or supports to the achievement of enhanced literacy and basic competency abilities. Finally the background questionnaire can obtain a detailed profile of skill demands placed on the respondent in his/her job across a broad range of competencies. Information from the background questionnaire is also used in assessing the stability of performance of the test over population subgroups and to improve the precision of the estimates themselves through conditioning.

Important themes treated in the common background questionnaire include:

- (1) Demographic information: age, sex, mother tongue, second language, family situation
- (2) Work force participation: occupation, labour force status and experience, income, attitudes toward work
- (3) Education and training: formal, on-the-job, informal
- (4) Literacy activities: reading/writing activities (job, community, home)
- (5) A profile of the skill demands of the workplace

ALL will follow the same development path as for the IALS background questionnaires. IALS background questionnaires were constructed through the same careful development, review, pilot testing, and quality control checks as the simulation tasks. In many cases, the challenge of creating valid and comparable background questions was even more difficult than developing achievement tasks. For example, providing accurate and parallel descriptions of occupations,

educational background, vocational training, and on-the-job training experiences required a careful analysis of each of the participating countries. Only variables that have the same meaning across all participating countries were used in comparative analyses.

The IALS background questionnaire was administered to each respondent prior to the cognitive assessment. Questions in the background questionnaire were divided into two groups - a group of questions which all participants were required to carry in the form specified (or in a form which demonstrably provides equivalent information) and a group of optional questions which countries were encouraged, but not obliged, to carry in the prescribed form.

A list of variables and a copy of the current draft of the international "model" background questionnaire has been included in the Frameworks: Working Drafts paper. The International Expert Committee has argued strongly to limit the amount of inter-country variation in questionnaire content.

On average the "mandatory" background variables required 15 minutes to complete in IALS. Administration of the optional background questions added roughly 10 minutes, for an average total completion time of 25 minutes.

The ALL background questionnaire will contain two supplementary sections for the two skill domains not directly tested through cognitive assessments. They are attitudes to teamwork and use of information and communication technology. The items for these two skill domains will be developed using the same development model as for the items in the cognitive assessments. This process is described in the next section. These items will also be tested in the pilot survey.

The ALL background questionnaire should require a total of 30 minutes of administration time: 10 minutes for the administration of the background variables, and 10 minutes for each of the two skill domains. National Study Teams will be required to review the draft background questionnaire with a view to :

Improving the content,

Identifying additional questions of national significance

Reaching international consensus on the questionnaire by February, 2000.

iii) Simulation Tasks

The ALL enters its item piloting phase with a draft framework describing each assessment area. The frameworks include:

- a definition to serve as a guide for what the test should measure and how scores should be used,
- a set of task characteristics that can be used in item development, and
- a set of variables that can be used to describe each item with respect to the task characteristics.
- A proposal on how performance can be summarised into levels for reporting purposes

ALL assessment development is well advanced. Frameworks for the ALL skill areas (prose and document literacy, numeracy and analytic reasoning) exist and have been used to develop item pools that have been successfully piloted in several countries. National Study Managers will be invited to circulate these documents among experts in their countries and to submit comments on them by December 30, 1999.

In addition, countries will be invited to send experts in the field of assessment to join item development teams that will work together through the first half of 2000 to develop the balance of items required for the pilot survey. This work will be accomplished both through a series of meetings with the framework writers and item development taking place in the members' home countries. At the first meeting, the frameworks will be explained in sufficient detail to allow national experts to develop items. The second meeting will provide the opportunity for team members to bring potential materials for task prompts and to begin the item development process together. This will be followed by item development in participating countries. The purpose then of the third meeting will be to review the items and to select those for use in the item pilot. After the third meeting, the selected items will be placed into the actual assessment instruments.

6.2 Data Collection Procedures

The comparability of survey results in ALL-type direct assessment depends upon standardised administration procedures. For IALS, a common administration guide was developed that included a script to be used by all interviewers while administering the assessment. The script provided standardised instructions for treating each interviewee with respect and sensitivity and for handling specific issues, such as:

- respondents who are hostile;
- respondents who quit midway through the assessment;

-
- respondents who have difficulty understanding the interviewer and the assessment instrument;
 - interruptions by other household members, telephone, or visitors;
 - extremely slow respondents; and so forth.
 - Provisions were made for documenting any administration irregularities.

Within each participating country, test administration was the responsibility of an experienced survey research organisation (governmental or private) that trained survey administrators and conducted standardised quality control checks of the field work. National Study Directors met to agree upon common training, data collection, and quality control procedures. Sample interviewer guides, record-keeping forms, and training materials were provided to all participating countries as models. Each participant was free to adapt these materials to reflect local operational realities but were obliged to seek prior approval from the International Study Team.

Many of the problems of comparability and possible bias in the IALS data can be traced back to a failure by countries to adhere to the agreed to procedures for administration. As a result, considerable more attention will be devoted in ALL to assuring that standards are met. Many of these requirements and quality assurance processes and procedures are set out in the draft Quality Assurance Specifications for the study. Note particularly that the Quality Assurance Specifications set out strict requirements for acceptable data collection agencies.

6.3 Pilot Surveys

For ALL, piloting will occur in two distinct phases. The first phase, an already completed item pilot, had two objectives:

- (1) validate psychometrically the items to be included in the ALL assessments and survey instruments
- (2) ascertain the appropriateness of the items to be included in the background questionnaire.

Item pilots have been conducted for numeracy, problem solving/analytic reasoning, teamwork and practical cognition. The pilots for numeracy and problem solving have been judged to be successful, whereas the item pilots for team work and practical cognition have identified a range of measurement problems with these scales. This aspect is critical to validate the psychometric properties of the assessments internationally.

The second phase is called the full pilot survey and will essentially be a small-scale dress rehearsal for the main survey. The full pilot survey is planned for the first quarter of 2001. It is

recommended that:

all ALL participants be required to conduct a 1,200 case full pilot survey

Such a pilot survey serves two purposes. Firstly, it provides empirical evidence to assess the performance of the test items. Inevitably some errors will be introduced in adaption to other languages. Items found not to be performing as expected in the pilot can be examined and corrected. It is not necessary that the pilot sample be truly representative to realise this goal. Countries are asked to field pilot samples that reflect the entire ability distribution.

Secondly, a pilot of this size affords participants exposure to the idiosyncrasies of administering household-based educational assessments of adults. Collection of this type of data is not trivial. To begin with most respondents will not have had prior experience with the government showing up at their door asking them to take a test. Special collection procedures and training are a must if non-response bias is to be contained. In addition, the application of a Balanced Incomplete Block assessment design introduces operational complexities well beyond those normally confronted in household survey research. Our intention is to use the ALL full pilot survey as a “dress rehearsal” for the main collection. Thus, participants should plan to test all production activities, including coding, editing, intra-country re-score and creation of an international data file. The only step that will not be performed is the inter-country re-score.

6.4 Data Collection For ALL

The outcome of the ALL item pilot survey will be a set of cognitive tests, organised into 30 booklets (as currently proposed), and a background questionnaire, which will incorporate items for measuring use of information and communication technology and attitudes to teamwork.

ALL participants will be provided with a machine readable version of the ALL questionnaires for review. They will then be required to send a representative(s) to a working meeting where possible modifications/additions/deletions will be discussed. A duly revised version will be circulated for final review and approval.

Once approved, ALL participants will be obliged to translate the questionnaires and provide a final copy to Statistics Canada for review. The object of this review will be to detect any deviation from the prescribed English language model. Countries will be required to revise any questions where the original intent of the question seems to have been compromised. Finally, countries will be required to forward an English version code book of their final data file.

In order to contain the risk implied in deviation from prescribed data collection procedures, Statistics Canada will conduct an audit of collection procedures prior to the commencement of main collection. This audit will necessarily include direct contact with the agency responsible for data collection.

One other apparently minor condition of participation are set out below. It is proposed that no country shall be allowed to pay a significant incentive to respondents. Experience suggests that this procedure has the potential to introduce uncorrectable bias into study results.

7 DATA PROCESSING

7.1 Database development

All ALL countries will be responsible for creating and validating data files following standard formats. All test results should be key entered with 100 percent verification. Background questionnaires will be subjected to statistical verification to guarantee acceptable average outgoing quality levels. All data files will be checked to ensure that each record has a unique identification number and that all data fall within expected ranges. Participants will be obliged to correct any anomalies so identified. Detailed guidelines will be provided by Statistics Canada.

7.2 Scoring

All ALL questions should be open-ended. Most questions will require simple correct-incorrect scoring. A few items will require more complex scoring of a short paragraph response. Countries should budget 15 minutes per case for scoring.

ALL participants will be provided with standardised scoring rubrics to inform the scoring process. By rendering all elements of scoring explicit these rubrics will serve to minimise the amount of variance between individual scorers. A team of chief readers from each country will meet to reach agreement on scoring guides. The chief readers will then direct scoring in their own countries.

The reliability of scoring within each country will be checked by having a second scorer re-score 20 percent of the booklets. The goal of the re-score within countries is to improve scoring to the point where inter-rater agreement approaches 100%. It is important to note, however, that participants will not be required to reconcile individual cases. Experience in the US indicates that average inter-rater agreement across all items exceeds 95%.

As an additional quality control check, a random sample of roughly 10% of test booklets will be selected and submitted to independent scoring. The objective of this re-score is to detect and correct systematic variation in the application of the standard scoring rubrics. More detailed scoring guidelines will be provided prior to implementation of the pilot survey and National Study Teams will be required to send a Chief Scorer to a training session immediately before scoring commences.

7.3 Statistical Analysis

IRT scaling. Item response theory (IRT) is utilized to estimate performance on each of the ALL skill domains. Since each respondent answers only a subset of the assessment tasks, the analysis takes into account responses to background questions along with achievement tasks to make estimates of performance (a technique known as "conditioning"). IRT assumes consistent patterns of which tasks are easy and which are hard, for all the tasks over all participating

countries. Preliminary analyses are required to verify that these patterns hold across participating countries. The steps involved in these type of analyses are described in detail in ETS's *IAEP Technical Report: Volume Two* and in *Adult Literacy in OECD Countries: Technical Report of the First International Adult Literacy Survey*.

Checks of validity and comparability. An important data analysis task is to check the validity of the results both within and across participating countries. A number of techniques are used. Traditional item statistics are used to identify initial problems in comparability, especially those that might occur due to printing or translation errors. Factor analysis is used to check for the unidimensionality of proposed scales. Within each country, IRT scaling is compared with across-country scaling in order to check that measures are comparable among countries or language groups. Differential item functioning (DIF) techniques are used to see if any questions are operating differently (e.g., they are unusually hard or easy, compared to other items) across participating countries. Tasks that are not operating in the same way in all participating countries are not included in summary statistics. Strategies for conducting this type of analyses are described in detail in the *IAEP Technical Report* produced by ETS.

Performance levels. Additional data analysis techniques will be applied to describe the nature of the results both across and within participating countries. In IALS results were reported by various performance levels. A number of methods are available for defining and reporting performance levels; most make use of a combination of expert judgment and statistical analysis of results. For example, in the LSUDA, four levels of reading skills were defined *a priori* based on a theoretical model of literacy. Readers wishing to review the process by which such levels are derived are referred to the NALS Technical Report. Tasks were then analysed and assigned to those levels. After the assessment was completed, actual performance data were used to confirm the theory and to provide actual numerical ranges for the levels. Results were then reported in terms of those levels, for example, the percentage of respondents performing at level one, and levels were described in terms of what individuals at that level were able to do. A similar level approach was used for reporting IALS results.

Relational analysis. In addition to reporting achievement levels for selected demographics and background characteristics, relational data will be presented as simple cross tabulations of results. In addition, for example, the performance levels of males and females by educational level. Multivariate methods of analysis, which require hypotheses to be carefully delineated and tested against the database, are employed for secondary analysis.

Sampling error and tests of significance. Since all results are derived from statistical samples and are, therefore, subject to sampling error, appropriate standard errors of estimates will be computed for all statistics. Likewise, the comparison of results between subgroups within a country or across countries are subject to sampling error, so statistical tests of significance will be applied to all comparisons. In the case of multiple group comparisons, adjustments will be made to account for the increased probability of identifying differences that are spurious, for example by using the Bonferonni adjustment for multiple group comparisons.

7.4 Data Processing for ALL

Statistics Canada will provide participants with a standard code book into which their data must be fit. Participants will be required to follow the specifications for creating these data files and ensuring their compliance to the prescribed format as outlined above. Where background questionnaires are captured, participants will invoke statistical quality assurance procedures to ensure outgoing quality levels of 99.5%. Countries will be required to key enter test results with 100% verification. Statistics Canada will vet data files from the pilot surveys and the main survey to ensure compliance. Countries will be obliged to correct any anomalies detected during the course of these analyses.

Countries will also be required to re-score and capture roughly 20% of all test booklets. A large fraction of this sample will be selected from early returns to inform and correct the scoring process. Inter-rater reliabilities of 95%+ are expected for the main study. For the main study, countries will also be required to independently re-score a random sample of an additional 300 cases, most likely from another country sharing the same language. This step is required to detect and correct any inter-country scoring bias.

Statistics Canada will, in co-operation with ETS, produce a standardised scoring guide. ALL participants will be required to translate these rubrics and apply them without modification or adaptation. Statistics Canada will arrange a meeting of chief scorers from each country prior to the beginning of scoring.

7.5 Statistical Analysis for ALL

A critical element of IALS was the conduct of a linking study in the US to provide for an explicit statistical linkage of the IALS scales to the scales developed for the National Adult Literacy Survey (NALS). Operationally, this linkage was achieved by modifying the basic IALS assessment design to a 21 booklet design where in each booklet contained two IALS blocks and one NALS block. Using this information, ETS was able to construct one set of scales for both studies.

For ALL, data from participating countries will be placed on these same international scales using the same IRT technology. As for IALS, a variety of statistical techniques will be used to establish the validity and comparability of the ALL performance data.

ALL participants will be required to attend two meetings, one to review performance data from the full pilot survey and a second to review performance data from the main assessment.

Countries failing to submit their data files by the specified deadlines will be liable for any additional costs associated with scaling their data separately.

8 REPORTING AND DISSEMINATION

ALL will generate two types of outputs, analytic products and data products. Each of these is described briefly below.

8.1 Data Products

ALL will generate two data products including:

- (1) a master-file of ALL data with results scaled for international comparison;
- (2) a public use micro-data file of ALL data with results scaled for international comparison;

The master-file of ALL data is a machine-readable file of ALL unit records from each national study where common variables appear in a common position and format. Idiosyncratic variables for each country are presented in separate sections of each record.

The ALL public use micro-data file is a unit record file that excludes those master-file variables that would lead to the identification of an individual. Access to the public use micro-data files affords researchers the maximum flexibility in analysis of study findings.

For the first year after release, access to the ALL public use micro-data file will be restricted to those researchers who submit a research request to a review panel for approval. After one year, researchers will have unlimited access to this file. As a reward for participation, all countries will have access to the entire public use data file after it has been vetted for confidentiality. ALL participants will have access to their own master-file and public use micro-data from other participants.

8.2 Analytic Products

The only "official" international analytic outputs of IALS took the form of an international comparative reports, *Literacy, Economy and Society: Results of the First International Adult Literacy Survey*, which was released on December 8, 1995 and *Literacy Skills for the Knowledge Society: Further results of the International Adult Literacy Survey*, released in November, 1997. The results of the SIALS round of collection are scheduled for release in June, 2000. These reports focus on the similarities and differences in the literacy profiles of participating countries and their implications for national and international policy making. Copies of these reports can be made available on request.

These reports were jointly published by the OECD and Statistics Canada. Participants had the opportunity to vet the publication prior to release.

8.3 Reporting and Dissemination for ALL

It is recommended that ALL follow a similar strategy for reporting and disseminating survey results, including the broad dissemination of a public use micro-data file. There are significant barriers to the release of such a file in some countries, but believe the benefits associated with release warrant the effort. It should be noted that the ALL international comparative report will only be released after obtaining appropriate reviews within each country, and participant consensus on required changes. Any country will have the option of withholding its data from the international report if it so chooses. This decision must be taken, however, prior to the data being scaled by ETS. Statistics Canada reserves the right to qualify national estimates in those cases where the evidence suggests deviations from design specifications have imported “significant bias”. National Study Managers will be asked to reach consensus with respect to minimal quality thresholds to be met and associated sanctions for non-compliance.

9. QUALITY ASSURANCE CONSIDERATIONS

9.1 Introduction

The set of quality assurance specifications produced for the Adult Literacy and Lifeskills survey (ALL) provides the minimum standards that must be met by each participating country (see The Adult Literacy and Lifeskills Survey Quality Assurance Specifications, September 23, 1999). These specifications are intended to eliminate any unnecessary variation in the survey structure and process between countries.

These standards are essential in order to achieve a high level of data quality, which will allow meaningful analysis of the survey results. Secondly, they will ensure that international comparisons of the survey estimates are accurate and reliable.

The ALL survey process in each of the participating countries will be monitored by the ALL Quality Assurance Team to ensure that each country meets these minimum requirements for data quality. The ALL QA Team will be reviewing the survey plans before the survey goes into the field, the processes occurring during the survey operation and as well the survey documentation showing the results of the survey process.

9.2 Considerations for the National Planning Report

The following areas need to be included in the National Planning Report and costed out when drawing up an ALL survey national budget:

-
- 1. Major analytic objectives of the ALL study;**
 - 1.1 Special subgroups in ALL;
 - 2. Target population;**
 - 2.1 Definition of the target population;
 - 2.2 Definition of exclusions from the ALL target population;
 - 2.3 Special additions to the target population;
 - 3. Method of data collection;**
 - 4. Sampling frame;**
 - 4.1 Description of the frame;
 - 4.2 Frame type;
 - 4.3 Data items on the frame for each stage of sampling;
 - 4.5 Quality assurance procedures;
 - 5. Sample size;**
 - 5.1 ALL target sample sizes;
 - 5.2 Special additions to the ALL sample size;
 - 5.3 Alternative methods to assure achieving the required sample sizes;
 - 6. Sample design;**
 - 6.1 Probability sampling;
 - 6.2 Multi-stage sample design considerations;
 - 6.3 Stratification;
 - 6.4 Primary sampling units (PSUs);
 - 6.5 Secondary sampling units (SSUs);
 - 6.6 Additional stages of sampling;
 - 6.7 Sample design derivation;
 - 7. Sample selection procedures;**
 - 7.1 Operational feasibility of sample design and selection;
 - 7.2 Time schedule for conducting the selection steps;
 - 7.3 Potential undercoverage of some population subgroups in data collection;
 - 7.4 Preparation of the sampling frame for selection;
 - 7.5 Computation of MOS and MIN MOS;
 - 7.6 Features of the selected sample;
 - 7.7 Output File;
 - 7.8 Quality assurance checks.
 - 7.9 Sample design derivation;
 - 8. Weighting;**
 - 8.1 Calculation of population weights;
 - 8.2 Calculation of replicate weights for the Jackknife variance estimation method
 - 8.3 Undercoverage analysis;
 - 8.4 Collection of variables used in benchmarking;
 - 8.5 Benchmarking to external population counts;

9.3 Team Composition

In general, each ALL national team should be made up of experienced, knowledgeable personnel, each with expertise in survey work in one of the fields of methodology, data processing, computer programming and data analysis. Further, expertise in coding levels of education and industry and occupation data to international standards is required.

The ALL survey team should consist of individuals (or institutes) which are capable of meeting the Quality Assurance (QA) specifications in each of these areas. National Study Managers are encouraged to use the QA Specifications as part of the terms of reference to ensure that the persons hired for each step of the survey process are aware of and capable of meeting the ALL survey standards. It is highly recommended that the same survey team be used for both the pilot and the main surveys.

9.4 Expert Meetings

As well as the regular meetings for the ALL National Study Managers, meetings with the team members from the individual fields of expertise will be conducted. These meetings will be scheduled so that they piggy-back on the National Study Manager meetings, however, a representative with knowledge in the areas of methodology, data processing, and scoring will each be required to attend two such meetings – one before the pilot survey and one before the main survey. These will be technical meetings where survey plans and requirements for a particular domain can be discussed in detail. Thus, travel required for these personnel to these meetings should be included in the overall survey costs.

9.5 Survey Instruments

Each ALL national project team will require translation facilities for the Background Questionnaire and test booklets. Note: several other master documents (eg. Scoring Guide, Interviewer Training Manual, Interviewer's Manual, etc) will also require translation from English. The cognitive items will need to be adjusted for each country according to the ALL item adaptation guidelines. Access to publishing software which is able to handle graphics and high-quality print is required.

Focus Group testing or other cognitive testing methods must be used to assess each countries' instruments before the Pilot test phase.

9.6 Sampling and Weighting

The person or persons who will provide the ALL statistical methodology will be responsible for providing the following items for the main ALL survey. It is essential that the sampling method used for the ALL be a probability sample design. The documentation for each item will need to be written in English, in a timely manner, and sent to Statistics Canada for review. The

responsible person must be available to answer questions which arise during the documentation review.

It should be noted that the pilot survey has fewer methodological requirements than the main survey (eg. Weights are not required). However, it is recommended that the same survey frame be used for the pilot survey as for the main survey in order to assess frame problems and response rates.

9.7 Data Collection

The data collection agency used for the ALL must have extensive national survey experience. The number of interviewers hired and their workload must conform to the ALL norms. Likewise, the payment scheme for interviewers should be such that response cases are maximized. Further, the interviewers must participate in an interviewer training session designed specifically for the ALL. The same data collection agency should be used for both the pilot and main surveys.

It is essential that the interviewing work be monitored (10% of cases) for the correct selection of one respondent per household (where applicable) and for the correct administration of the questionnaire and test booklets, as per the QA Specifications.

With a required 70% response rate over all stages of sample selection required for the main survey, it is essential that contact protocol and follow-up procedures for non-response cases be built into the interviewer process. Thus, documentation of survey responses must be carried out throughout the collection activity, after each 20% of cases are completed.

9.8 Data Capture, Editing and Coding

The QA specifications for data capture, editing and coding should all be adhered to for both the pilot and the main surveys.

For paper and pencil instruments, a manual data capture system must be used to capture the ALL Background Questionnaire and test booklet scores data. This system must be a 100% verification system, tested prior to the actual data capture operation. Computer-assisted systems can be used for the Background Questionnaire – again with testing being necessary.

Once the data have been captured, an automated editing system must be in place to check for valid ID numbers, establish final status codes for all records, perform range checks, conduct consistency edits and perform imputation for the failed edit records. Some manual intervention will be required to check question flows and that the correct populations flow through the appropriate sections.

Where codes need to be assigned on the ALL file (eg. ISCED, ISIC, ISOC, etc) a trained coding

team will be required. Automated coding can be used where an approved system already exists. The coding must be shown to have been performed with a maximum 6% error rate.

Finally, the data must be stored into the standard ALL record layout and vetted for confidentiality.

9.9 Test Scoring

A scoring team for the cognitive tests is required with between 5 to 10 persons (such that no scorer scores more than 750 booklets). These scorers must have been trained in this operation – with the head scorer having attended an international scoring meeting.

During the pilot test, a thorough analysis of the responses obtained for each question must be performed by each country. In this way, unexpected answers or answers unique to a country can be identified and included in the Scoring Manual for the main survey.

For quality assurance purposes, the ALL cognitive tests will require two scoring operations: an intra-country rescore (within country) and an inter-country rescore (between two countries). Note that only the intra-country rescore is required for the Pilot test. Scorers should have multi-lingual capacities so that an inter-country rescore can effectively be carried out. For both these rescoring operations, monitoring of the errors, correction of the errors and then documentation of the process must be performed. Both the intra-country and inter-country rescore operations are extremely time-consuming operations since the correction of errors can involve a complete rescore of the tests.

For countries who participated in the IALS and would like a comparison of literacy skill estimates between the two surveys, an additional scoring of booklets from the IALS survey will need to be done. This scoring will determine if there are any drifts in the scoring standard between the two surveys. About 300-500 IALS booklets would need to be rescored for this operation.

10 MANAGEMENT, STAFFING AND SCHEDULE

The following section of this report outlines a proposed national and international management structure and a tentative implementation schedule. For convenience, a summary of proposed roles and responsibilities of the various partners has been included as Appendix A.

10.1 International Management and Advisory Structure

Experience suggests the need for a strong, highly visible management and oversight function in any statistical study that purports to generate internationally comparable data.

Such a role becomes critically important in a study such as ALL that must maintain psychometric equivalence across countries. Even the most robust of psychometric undertakings are frighteningly susceptible to subtle changes in presentation and administration. It is proposed that Statistics Canada assume primary responsibility for overall project management for ALL. They would be responsible for arranging for international meetings, maintaining communication with participating countries, co-ordination of technical consultants, establishing project budgets and schedules and quality assurance at every stage of the project.

T. Scott Murray of Statistics Canada and Marilyn Binkley of NCES will share the role of international study directors. Both have extensive national and international experience related to assessment of performance.

The Educational Testing Service has been retained to advise on the design of the assessment, to analyse the test data from the pilot surveys, to scale and link the final ALL assessment data to the NALS/IALS scales and to produce a comprehensive set of tabular output complete with standard errors. ETS performed this function for IALS superbly. Irwin Kirsch, Kentaro Yamamoto and Don Rock will assume primary responsibility for psychometric issues. Yvan Clermont of Statistics Canada will assume the role of Manager of International Operations.

Westat will assume primary responsibility for background questionnaire development. Statistics Canada will assume primary responsibility for quality control of sampling, test adaptation, data collection, scoring and some aspects of database development. It is recommended that Statistics Canada and the OECD manage the reporting and dissemination function with input from ETS. Trevor Williams is managing this activity.

A Project Advisory Group (PAG) has been established for the ALL to provide strategic advice and operational oversight for the project. Each participating country automatically has membership on the PAG, although membership is not limited to participants. The OECD is also represented on the PAG. The PAG has met twice, most recently in Paris in June 1999.

An International Expert Committee has been established to provide advice with respect to quality

assurance in ALL and to review evidence related to comparability, validity and reliability prior to publication of ALL results. This Committee will function at arms length to the International Management Team. Participants will be asked to nominate experts to sit on this team. Authority to place experts on the Committee will, however, rest with the ALL Project Advisory Group. Lars Lyberg of Statistics Sweden will chair the expert Committee.

It should be noted that ALL, like IALS, will be managed on a collegial model where all partners will have input into key design decisions. All such decisions will be arrived at by consensus of participating countries.

10.2 National Management and Advisory Structure

Each participating country will appoint a national project director who will be responsible for the overall management and quality of the assessment. It is also recommended that participating countries involved national experts familiar with testing in general and IRT estimation methodologies. Note that the Independent Quality Advisory Committee established for ALL has recommended that National Study Managers possess broad experience in the area of both assessment and household surveys.

Each country will be encouraged to assemble an advisory group that includes representatives from the important literacy advocacy, policy making, and funding groups. Such a group was created for the IALS study in several countries and was deemed to play an important role in disseminating the study results.

10.3 Project Schedule, Tasks and Deliverables

The time period for the project is September 1, 1999 through December 30, 2000. The following provides a schedule of tasks, meetings, and deliverables.

Schedule of Tasks	Completion Date
<u>Development</u>	
First Meeting of National Study Managers	1999/09
Framework Review	1999/09-1999/12
Item Development for Prose, Document, Numeracy	1999/09-2000/06
<u>Pilot Survey</u>	
Translation/Adaptation of Items	2002/02-2000/07
Pilot Advance Design Meeting	2000/03
National Planning Report Submitted	2000/05
Decision Point for Participation	2000/06
Pilot Survey	2001/06
Pilot Data to Statistics Canada	2001/06

Pilot Data Analysis	2001/06-2001/08
Pilot Review Meeting	2001/09
<u>Main Survey</u>	
Main Advance Design Meeting	2001/10
Main Scoring Meeting	2002/01
Main Collection	2000/02-2002/01-2002/04
Data Files to Statistics Canada	2002/08
Data Files to ETS	2002/09
Scaling, Linking and Analysis	2002/09-2002/12
Main Assessment Review	2003/01
Final Ability Estimates Available	2003/02
Draft International Comparative Report	2003/08
Release of International Comparative Report	2003/12
Release of International Micro-data	2001/09

Schedule of Meetings	Date
-----------------------------	-------------

Development

First meeting of NSM	1999/09
Item Development	1999/10
Item Development Review	2000/03

Pilot Survey

Pilot Advanced Design Meeting	2000/03
Pilot Scoring Meeting	2001/01
Pilot Survey Review Meeting	2001/09

Main Survey

Main Advanced Design Meeting	2001/10
Main Scoring	2002/03
Main Assessment Review	2003/01
International Comparative Report Review	2003/08
International Comparative Report Release	2003/12

Appropriate representatives from Statistics Canada and ETS will attend each of the proposed meetings.

Countries are required to send appropriate personnel to each of the meetings listed above. Although not required, National Study Managers are encouraged to attend all meetings. Countries should also budget to host at least one of these meetings. The costs associated with hosting a meeting are minimal, at a maximum US \$5,000.

Countries should also budget for establishing an Internet connection where one is not already available. Experience with IALS suggests this type of communication is indispensable to the success of the study.

11. BUDGET

Each participating country will be responsible for all within-country project costs and for sending representatives to international meetings. As mentioned above, they should also budget a modest amount to cover incidental expenses associated with hosting a meeting.

Due to the nature of the project, ALL will incur significant overheads. These overheads cover management, design and statistical analysis throughout the project. A summary of these expenditures by project phase is shown below. The fiscal years shown run from April 1 to March 31.

International overheads in this phase of the project relate to general management costs and to the retention of experts to lead the development of each of the survey components i.e.

Teamwork	David Baker, AIR
Problem solving	Eckhard Klieme, Max Planck
Numeracy	Iddo Gal, University of Haifi
Prose Literacy	Irwin Kirsch, ETS
Document Literacy	Irwin Kirsch, ETS
Background questionnaire	Trevor Williams, Westat
Practical Cognition	Robert Sternberg, Yale University
Information and Communication	
Technology	Jean-Paul Reeffer, Ministère de l'Éducation Luxembourg

Additional consultants may be required. Funds are also required to support the involvement of ETS as technical consultant on the design of the study and the assessments. Participants will also be responsible for financing their own national item pilot and the work of any experts who they involve in the review of the frameworks and the development of the item pools or the background questionnaire.

It is recommended that a similar practice be followed for the full pilot survey and the main data collection in ALL. In SIALS, these international overheads totalled roughly \$750,000 US for ten countries. Because Canada and the US subsidized the first IALS round heavily, these costs appear to be considerably greater than the \$15,000 and \$25,000 US which countries were asked to contribute. It is our judgement that the need to contribute financially to the undertaking has a salutary impact on countries' commitment to the study far beyond the amount of money involved. Statistics Canada will continue to seek funding from international agencies to offset some of these costs. If such funding materializes, countries will be collectively offered the choice of a rebate or extension of the analytic output. In practical terms, this means that ALL

participants will be required to contribute roughly \$300,000 US towards international overheads for the pilot and the main study over a three or four year period.

ALL International Overheads (in \$US)						
Per Country						
Activity	Fiscal Year					Total
	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	
ALL Management Group		\$4,500	\$4,000	\$4,000		\$12,500
BQ Development Work		\$10,000				\$10,000
Assessment Development Work		\$60,000	\$10,000			\$70,000
Assessment Processing			\$35,000	\$85,000		\$120,000
Quality Assurance Team		\$12,500	\$12,500	\$12,500		\$37,500
International and Technical Reports					\$50,000	\$50,000
Total		\$87,000	\$61,500	\$101,500	\$50,000	\$300,000

12. NEXT STEPS

Following the first meeting of National Study Managers prospective participants are asked to:

a) begin the process of drafting a National Planning Report

This report will serve to inform national deliberations concerning the cost and operational implications of participation and provides the first element of the ALL quality assurance strategy.

Production of a National Planning Report of sufficient quality to meet this latter need requires the presence of National Study Manager (or suitable proxy) and a supporting multidisciplinary national study team. The National Planning Report can also serve as a useful guide where participants are required to tender the actual execution of the study, one that ensures that prospective bidders are aware and capable of meeting the study's operational and technical requirements. Countries are asked to submit draft reports to Statistics Canada by February 2000 with a view to having final proposals by June, 2000.

b) contribute national expertise to the development of the item pools for prose literacy, document literacy and numeracy

Given the fact that the ALL instruments are likely to see wide spread use it is important that the item pools be broadly representative culturally, linguistically and geographically. Prospective participants are asked to consider the contribution of national expertise to the development of the remaining items needed for the pilot survey. This will ensure the needed diversity and provide first hand exposure to the pragmatics of population assessment, exposure that has proven useful in presenting national results. This activity will continue through June, 2000.

c) contribute to development of the background questionnaire

The ALL background questionnaire serves an obvious and important role in the analysis and interpretation of study results. The International Expert Committee has strongly recommended that inter-country variation in questionnaire content be kept at an absolute minimum to guard against the introduction of bias. At the same time the currently proposed background questionnaire exceeds the national time available. Prospective participants are asked to identify an individual to assist the International Study team in finalizing the ALL background questionnaire. This activity will continue through March, 2000.

d) have the draft framework documents for each skill domain received by national experts

The scientific credibility of the entire ALL exercise rests on the strength of theoretical frameworks upon which the items will be built. It is important, therefore, that prospective participants understand and contribute to the current drafts. Such understanding is crucial to interpreting and explaining national results. Countries are asked to subject the current drafts to expert review with a view to providing written comments by December, 1999.

Collectively these activities will permit countries to formally commit to the ALL study by no later than June, 2000, the date by which the instrumentation will be finalized and the deadline for submission of final vetted National Planning Reports. This will allow the pilot survey to be fielded as planned in the first quarter of 2001 and the main survey to be conducted in the first quarter of 2002.

APPENDIX A

ROLES AND RESPONSIBILITIES

APPENDIX A

THE ADULT LITERACY AND LIFESKILLS SURVEY Responsibilities of Statistics Canada

Statistics Canada will provide:

- interviewer training guide and workbooks
- interviewer manuals
- collection procedures manuals
- a model screening questionnaire
- interviewer task administration guides
- observation and scoring forms
- scoring guides

Statistics Canada will also:

- specify format and content for items for inclusion on background questionnaire
- specify a suggested content and format for common optional questions which may be carried on the background questionnaire
- provide an assessment design, specifying item order within blocks and block order within booklets
- review proposed sampling schemes
- graphics files, in English, for final test items and the newspaper. These will be provided in Adobe Illustrator/Pagemaker

Countries will:

- contribute \$300,000 US towards the costs of overall project management
- adhere to all design specifications in implementing the study
- provide a planning report to Statistics Canada for review and approval
- provide a back translation of non-English versions of the assessment and background questionnaire
- re-score a minimum of 20% of booklets
- re-score up to 300-500 booklets from another country
- subject the background questionnaire to cognitive review
- provide copies of all collection manuals, procedures and related forms
- submit to a pre-collection implementation audit
- collect the data following the prescribed procedures
- provide a clean data file in a prescribed format
- provide technical documentation in a prescribed format in English and, wherever possible, in machine readable format
- provide code books in English, for the pilot and main assessment data files
- submit copies of final questionnaires and assessment materials to Statistics Canada for approval to print prior to the pilot and the main assessment
- send appropriate representatives to all planned study meetings
- agree to host at least one of the planned study meetings

- guarantee that clean data files from the pilot and main assessment are sent to Statistics Canada on agreed to dates
- allow a public use micro-data file to be released within one year of release of the international comparative report
- agree to refrain from any publication or other dissemination of the assessment booklets without prior written authorisation of ETS
- agree to refrain from any publication or other release of data from the study prior to release of the international comparative report

Responsibilities of the Educational Testing Service

The Educational Testing Service will assume primary responsibility for:

- briefing of prospective participants on the technical aspects of the assessment
- the statistical analysis and scaling of pilot data and related modifications to items
- providing recommendations for final design
- the statistical analysis and scaling of main assessment data
- the provision of univariate estimates of proportion of the population at each level on each scale, performance means and related standard errors
- the provision of general technical advice to Statistics Canada throughout the study.

APPENDIX B

INTERNATIONAL PARTNERSHIPS AND FRAMEWORK DEVELOPMENT TEAMS

Appendix B: International Partnerships and Framework Development Teams

One of the main challenges associated with developing a large-scale survey in an international context, such as the Adult Literacy and Lifeskills survey, is ensuring that an international perspective is maintained throughout survey and item development. ALL has made an effort to ensure the inclusion of multiple perspectives in scale development to date, and to avoid a dominantly North American perspective whenever possible. This has been done through staffing the framework development teams so that, ideally, they reflect multicultural expertise across the domains (see below for a complete listing of current teams).

To further expand international involvement, activities are planned along two fronts of the ALL. First, prospective participants are invited to submit nominations for additional experts for the prose, document and numeracy development teams as they prepare for the pilot survey in 2001. These experts will aid in ensuring that the finalization of the frameworks and the development of items reflect the multi-cultural development goals of ALL.

Development Teams

Prose and Document Literacy

- Irwin Kirsch, Educational Testing Service, U.S.
- Stan Jones, Statistics Canada, Canada
- Peter Mosenthal, Syracuse University, U.S.

Numeracy

- Marilyn Binkley, National Center for Education Statistics, U.S.
- Iddo Gal, University of Haifa, Israel
- Mieke van Groenestijn, Hogeschool van Utrecht, the Netherlands
- Stan Jones, Statistics Canada, Canada
- Myrna Manly, El Camino College, California, U.S.
- Yvan Clermont, Statistics Canada, Canada
- Laura Salganik, Education Statistics Services Institute, U.S.
- Mary Jane Schmitt, National Center for the Study of Adult Learning and Literacy, Harvard University, U.S.
- Dave Tout, Language Australia, Australia

Problem Solving/Analytic Reasoning

- Judith Ebach, Institute of Educational Research, IBF, Bonn, Germany
- Joachim Funke, University of Heidelberg, Germany
- Anne Hensgen, Institute of Educational Research, IBF, Bonn, Germany

-
- Eckhard Klieme, Max-Planck Institute for Educational Research, Berlin, Germany
 - Jean-Paul Reeffer, Ministry of Education, Luxembourg
 - Eftychia Sidiropoulou, Institute of Educational Research, IBF, Bonn, Germany

Practical Cognition

- Guillermo Gil, Spanish Ministry of Education, Spain
- Elena Grigorenko, Yale University, U.S.
- Robert Sternberg, Yale University, U.S.

Teamwork

- David Baker, American Institutes for Research, U.S.
- Michael Campion, Purdue University, U.S.
- Susan Cohen, Center for Effective Organizations, University of Southern
- Lisa Horvath, George Washington University, U.S.
- John Konstant, Education Statistics Services Institute, U.S.
- Walter McFarland, American Institutes for Research, U.S.
- Lynn Offerman, George Washington University, U.S.
- Eduardo Salas, Naval Air Warfare Center, U.S.
- Laura Salganik, Education Statistics Services Institute, U.S.

Information and Communication Technology

- Jean-Paul Reeffer, Ministry of Education Nationale, Luxembourg
- Graham Lowe, University of Alberta, Canada
- Julie McAuley, Statistics Canada, Canada
- Albert Tuijnman, University of Stockholm, Sweden

Background Questionnaire

- Albert Tuijnman, University of Stockholm, Sweden
- Anna Borkowsky, Swiss Federal Statistical Office Switzerland
- Patrice de Broucker, Statistics Canada, Canada
- Eveline Von Gaessler, German Ministry of Education, Germany
- Thomas Healy, Organization for Economic and Cooperative Development
- David Kastberg, Westat, U.S.
- Helmut Kuwan, Infratest Expert, Germany
- Trevor Williams, Westat, U.S.